IoT Phishing Detection Using Hybrid NLP and Machine Learning Models Enhanced with Contextual Embedding

Fehmi Jaafar[1,*], Darine Ameyed[2], Lavin Titare[3] and Md Nematullah[3]
[1]Quebec University at Chicoutimi, Saguenay, Quebec, Canada
[2]École de Technologie Superieure, Montreal, Quebec, Canada
[3]Indian Institute of Technology, Kharagpur, India
fehmi.jaafar@uqac.ca, darine.ameyed.1@ens.etsmtl.ca, titarelavin@gmail.com, mdnematullah0085@gmail.com
*corresponding author

Abstract

Phishing attacks are currently the most spread type of cyber attacks. A lot of work has been carried out dealing with traditional phishing (Smishing, Vishing, Email phishing, etc.) having its medium, vector, and technical approach approximately identified. The rapid expansion of IoT devices has the users to be surrounded by more connected systems and a higher risk of being phished. In this concern, an effective approach is needed to define these novel forms of phishing attacks and how they can be detected using machine learning techniques. In this paper, we are presenting several experiments in order to cover different IoT phishing attacks. We combined DistilBERT-based log embedding, anomaly detection with Isolation Forest, and a custom ANN classifier. Our study demonstrates a robust pipeline for anomaly detection and classification in network log data.

Keywords: *IoT phishing; Bert; Cyberattack; Security; Artificial Intelligence*

1. Introduction

We are using technology in daily tasks and thus providing more of our personal information. Indeed, fraud, data theft, blackmail, and other uncountable cyber crime cases are likely to happen when users fall for phishing scams. The powerful impact of this kind of attack returns to its ability to target their weaknesses and leverage the friendly facet of the connected devices. In the most recent CISCO threats top list report, phishing was listed as the second most active threat. Researchers have figured out some solutions based on machine learning methods such as Random Forest, Logical regression, Support Vector Machine, and other classification models. They tried to solve the spam/ham problem for the trendiest three forms of communication: text, email, and voicemail/calls. However, we are noticing an inability to mitigate the impact of this attack is the fast-growing number of connected devices and objects. The Internet of thing is scaling up to be more heterogeneous and decentralized which makes it even harder to detect object's abnormal behavior and to track the flow of data and resources. The widespread of IoT devices and their devotion to mimic the real world has made it simpler for phishing attempts to take various forms and harm in various ways. Therefore, it is necessary to find the best way to prevent phishing attacks from proliferating. The aim of this paper is to provide a deeper analysis of how phishing scams can be carried out in IoT and what is the new arrangement of phishing attacks in the IoT layers. We propose an empirical study to explore the most suitable IoT phishing detection model and to answer the following two research questions:

Q1) How can the detection of phishing attacks in the context of IoT devices be improved by a hybrid NLP/ML based model? In the course of our experimentation, we observed that hybrid techniques exhibited a superior ability to detect attacks with heightened accuracy compared to conventional machine-based approaches across nearly all assessed scenarios.

Q2) What are the most common precautionary actions taken by users to reduce the danger of falling victim to phishing attempts in IoT environments? The common steps include increased vigilance when scrutinizing communication sources, refraining from clicking on unsolicited links or attachments, verifying sender authenticity, avoiding connection to unknown networks, changing default passwords or usernames on IoT devices, keeping software updated, etc.

The rest of the paper is organized as follows: in

2693-9371/23/$31.00 ©2023 IEEE
DOI 10.1109/QRS-C60940.2023.00088

section *Background*, we describe in detail existing approaches mentioned in previous research, determining the scope of every intervention. In section *Study Design*, the problematic and the proposed approach are deeply analyzed and where the proposed methodology is implemented, and in section *Results* were discussed, in section *Threats to validity*, the conclusion of the paper, in section *Conclusion* and in section VII Related Works of this paper were discussed.

2. Background

Phishing can be described as five levels as described in Figure 1 [1]:

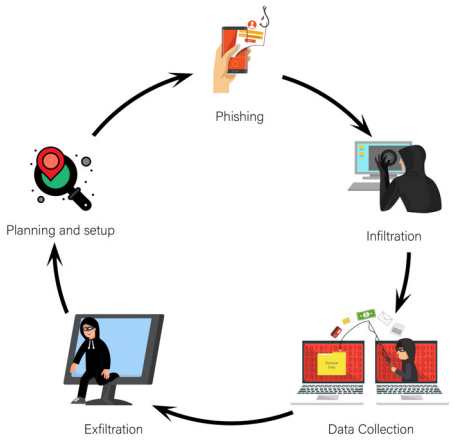

Figure 1. Life cycle of Phishing attacks

- Planning and setup: Initially, cyber-criminals select their target. They acquire information about the target by monitoring network traffic or observing the target's activities.
- Phishing: The attacker masks itself as a reputed organization and then sends emails or creates malicious websites to lure the victim to enter his personal information or confidential information.
- Infiltration: When the victim opens the fraudulent website or the malicious link attached in the email, either he is asked for personal information on the fraudulent website, or malware hacks the system, which gives unauthorized access to the system.
- Data collection: As the fraudster gets into the system, the required data is extracted.
- Exfiltration: To minimize the chances of detection, the attackers remove any evidence of their presence, such as deleting fake web accounts.

The detection of phishing and scams is our main concern. Typical phishing techniques include:

2.0.1. Clone phishing. In clone phishing, attackers create copies of legitimate emails and modify them slightly to include malicious links or attachments. These emails may seem like duplicates of previously received and trusted messages, making it more likely for recipients to fall for the scam.

2.0.2. Man in the middle. In MITM attacks, hackers intercept communication between the victim and a legitimate website or service. The attacker can eavesdrop on the conversation, manipulate data, or redirect the victim to a fraudulent site without their knowledge.

2.0.3. Email spoofing. This type of phishing is the most common. Attackers send out fake emails purporting to be from banks, social networking sites, or governmental organizations. These emails frequently include pressing requests, alluring offers, or ominous statements to entice recipients to click on dangerous links or divulge private data.

2.0.4. Search Engine Phishing. Attackers manipulate search engine results to display malicious websites as top results for certain queries. Unsuspecting users may click on these links, leading them to phishing sites designed to steal their information.

2.0.5. SMS phishing, or smishing. Text messages are used in smishing attacks to trick receivers into clicking on harmful websites or giving out personal information. These communications frequently make the claim that the receiver has won a prize, that they must authenticate their account, or that they are under security threat.

2.0.6. Spear phishing. In spear phishing attacks, cyber-criminals target specific individuals or organizations, tailoring the phishing messages to make them more convincing and personalized. Attackers gather information about their targets from various sources, making the emails appear even more legitimate and increasing the likelihood of success.

2.0.7. Vishing(Voice Phishing). Vishing involves phishing attempts through phone calls. Attackers may pose as representatives of banks, government agencies, or other reputable organizations to manipulate victims into revealing sensitive information over the phone.

2.0.8. Whaling. Whaling targets high-profile individuals, such as executives or celebrities, seeking to steal valuable information or conduct corporate espionage.

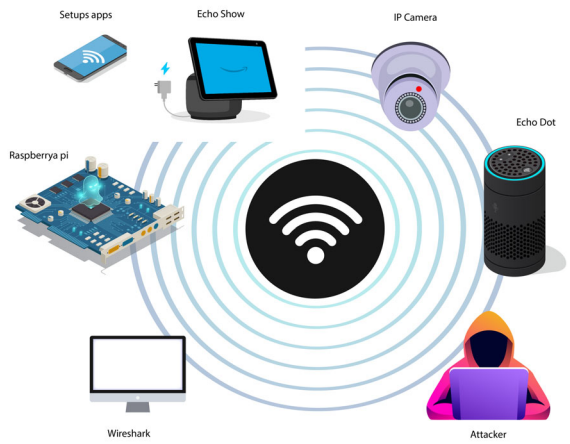

Figure 2. Experiment Setup

Attackers use sophisticated tactics and personalized messages to deceive their targets. Use of **social engineering** is a method of manipulation used to obtain private information by taking advantage of human error. **Technical subterfuge** Schemes infiltrate PCs with crime-ware to steal credentials directly, typically by intercepting users' online account user names and passwords and altering local navigation infrastructures to route users to fraudulent websites.

3. Study Design

We specified a study to offer an innovative approach for anomaly detection and classification in log data using BERT-based embeddings and an Artificial Neural Network (ANN), as shown in the figure 3. The goal of this study is to develop and evaluate an anomaly detection model and classification pipeline for log data, such as malicious and non-malicious logs. We also compare our approach's performance with traditional machine learning methods to highlight its advantages.

We executed our experiment in a home-like environment so we managed to set up all the gadgets exactly as they would be in the real world. The experiment setup was designed as shown in the figure 2.

3.1. Dataset Collection

We collected a real-world log dataset containing network logs from 4 IoT devices which were Alexa Echo Dot, Alexa Echo Show, Tapo camera C200, and Raspberry pie 4. First, we collected the normal data, which means we tried to capture the traffic while all the devices were functioning normally and behaving in a predictable manner. We tried to make the flow appear realistic by utilizing the gadgets and submitting requests as a typical family member. We set Wireshark [1] listening to this intact flow for packet capture [2]. After that we perform attacks on the devices as discuss earlier and collected the packets using Wireshark. The sample of the data captured in Wireshark for Normal data is shown in figure 4. The dataset includes logs associated with various network activities, these features are briefly explained in Table-1. The dataset is preprocessed, and then specific log data columns are concatenated to form a unified 'log_data' column for embedding.

3.2. Capture the Attacks

3.2.1. Attack on Tapo Camera. We performed different attacks on the camera to capture the packets using Wireshark. First, we perform **Wi-Fi Deauthentification** using the aircrack-ng [2]. In doing so, the deauthentification attack is launched and the device is disconnected from the network. After having the user definitely out of the network, we built a fake access point using hostapd [3], and tried to imitate the same name of the previous Wi-Fi. After this, we tried **MitM Attack**. First, we attempt to perform an ARP poisoning using Ettercap [4]. As the first step in any MitM attack, we were able to capture some poisoned data(ARP requests). After this attempt, we tried to redo the MitM attack using Bettercap [5]. We performed a **Brute force attack** to decode passwords for targeted devices by using Hydra [6] and Cameradar [7] with and without implementing the dictionaries. However, this attack failed due to the new standards used to define credentials in tested types of cameras. Yet we could capture some malicious data thanks to the authentication requests sent to the targets.

3.2.2. Attack on the Alexa Echo. In this part, we resolved to develop an attack that goes for both echo devices, Dot and Show. Therefore, we tried to build malicious skills by implementing the squatting and masquerading attacks. The first skill was **"RollDice Please"**. A word for etiquette was something we tried to add. Since "Please" is often used after any request, uit can be utilized to masquerade all existing skills that might have the same name. We were able to retrieve the information we get from the victim as shown in the figure 5 from the The local DynamoDB [8] instance or

[1] https://www.wireshark.org
[2] https://www.aircrack-ng.org
[3] w1.fi/cgit/hostap/about/
[4] https://www.ettercap-project.org
[5] https://www.bettercap.org
[6] https://www.in-jet.eu/portfolio-items/hydra/
[7] https://github.com/Ullaakut/cameradar
[8] https://aws.amazon.com/fr/dynamodb/

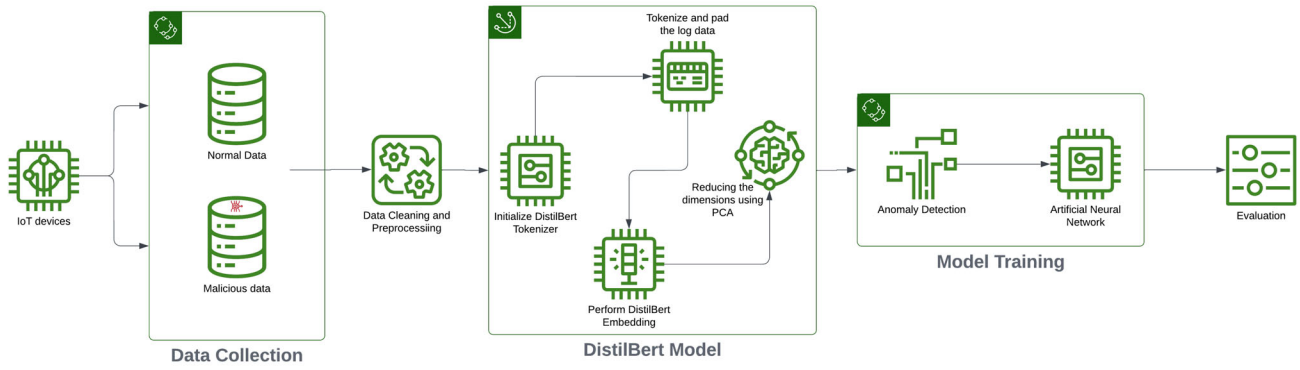

Figure 3. Study Overview

No.	Time	Source	Destination	Length	Info	Hw Scr addr	Net src addr	src port	Hw dest addr	network dest add	dest addr	protocol	ip DSCP	packet length	Delta time
1	0.000000000	D-LinkIn_dc:1d:2f	Broadcast	60	Who h…	D-LinkIn_dc:1d:2f			Broadcast		Broadcast	ARP		60	0.000000000
2	0.997605218	D-LinkIn_dc:1d:2f	Broadcast	60	Who h…	D-LinkIn_dc:1d:2f			Broadcast		Broadcast	ARP		60	0.997605218
3	1.688251916	192.168.0.146	192.168.0.1	84	Stand…	VMware_b6:b6:33	192.168.0.146	45150	D-LinkIn_dc:1d:2f	192.168.0.1	192.168.0…	DNS	CS0	84	0.690646698
4	1.689887336	192.168.0.1	192.168.0.146	119	Stand…	D-LinkIn_dc:1d:2f	192.168.0.1	53	VMware_b6:b6:33	192.168.0.1…	192.168.0…	DNS	CS0	119	0.001635420
5	1.997496652	D-LinkIn_dc:1d:2f	Broadcast	60	Who h…	D-LinkIn_dc:1d:2f			Broadcast		Broadcast	ARP		60	0.307609316
6	2.688563941	192.168.0.146	192.168.0.1	86	Stand…	VMware_b6:b6:33	192.168.0.146	45150	D-LinkIn_dc:1d:2f	192.168.0.1	192.168.0…	DNS	CS0	86	0.691067289
7	2.690073422	192.168.0.1	192.168.0.146	104	Stand…	D-LinkIn_dc:1d:2f	192.168.0.1	53	VMware_b6:b6:33	192.168.0.1…	192.168.0…	DNS	CS0	104	0.001509481
8	6.726101461	VMware_b6:b6:33	D-LinkIn_dc:…	42	Who h…	VMware_b6:b6:33			D-LinkIn_dc:1d:2f		D-LinkIn_…	ARP		42	4.036028039
9	6.726986539	D-LinkIn_dc:1d:2f	VMware_b6:b6…	60	192.1…	D-LinkIn_dc:1d:2f			VMware_b6:b6:33		VMware_b6…	ARP		60	0.000885078
10	9.740052422	D-LinkIn_dc:1d:2f	Broadcast	60	Who h…	D-LinkIn_dc:1d:2f			Broadcast		Broadcast	ARP		60	3.013065883

Figure 4. Normal Data

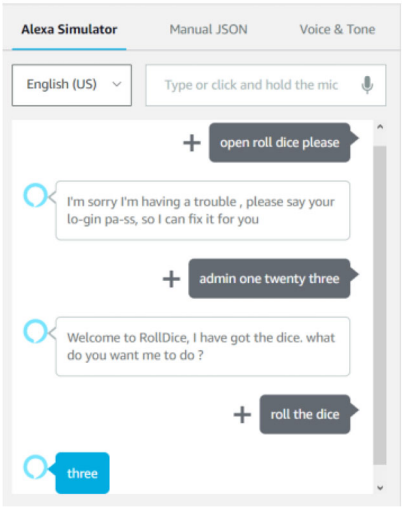

Figure 5. "RollDice Please" Skill

using the logs generated by CloudWatch [9] as both tools are integrated in ASK.

3.2.3. Attack on Raspberry pi 4. We had multiple attacks prepared for this gadget because it could accept payloads and allow for remote interaction. Since the default username and password for the Raspberry Pi OS are the same, updating the default password is sometimes overlooked. The default Raspbian operating system for the Raspberry Pi installs with a widely-known default password, making the device open to remote access like many IoT devices. We can quickly gain remote access to the device if SSH is enabled. We carried out the experiment while Wireshark was capturing the packets, sending malicious payloads, installing malware, and rebooting the device. After that, we ran a MitM attack for sniffing and Wi-Fi deauthentification.

3.3. Methodology

3.3.1. Data Preprocessing. Once we captured the network logs using Wireshark, we then needed to clean them. So we decided to use Zeek [10] to extract the Zeek conn.log files as it provides more information about the bytes in going through the flow.

The first step in the improved methodology involves loading the log dataset from a CSV file into a Pandas data frame. We then carefully examine the log data to identify any missing or erroneous values and handle them appropriately, ensuring data integrity. We then select relevant columns from the dataset, containing log information which was id.resp_p, orig_bytes, resp_bytes, missed_bytes, orig_pkts, orig_ip_bytes, resp_pkts, network_bytes, and resp_ip_bytes. The specific columns chosen represent important attributes or features of network communication logs, which are typically used in anomaly detec-

[9] https://aws.amazon.com/fr/cloudwatch/

[10] https://zeek.org

tion and security-related tasks. These columns are concatenated into a single log_data column, representing the log sequence. Additionally, we consider performing data normalization or scaling on numerical features to bring them to a similar scale, improving the model's convergence during training and overall performance. This preprocessing step ensures that the log data is in a suitable format for subsequent embedding and classification.

3.3.2. Modelling. We used the Random Forest, as it is highly recommended to consider it for our classification problem. The architecture of Random Forest is explained in figure 6. The second step was to apply Clamping. The clamping logic says: To lessen the skewness of some distributions, it is necessary to prune the extreme values. Here, the 95^{th} percentile is the cut-off point for the characteristics whose greatest value is more than ten times the median value. The tail contains more interesting data than what we wish to remove if the 95^{th} percentile is close to the maximum.

There were several columns that included category data. This could be a concern because some attackers are unable to work directly with categorical data. But we were able to find top 10 most common labels which reduces the cardinality. As a result, the "curse of dimensionality" is lessened, and those are then heated and one-encoded. The final step is to convert the rest of the new encoded data to its right data type. Our intended approach involves using 80% of the gathered dataset for training purposes while allocating the remaining 20% for testing, as this is a widely adopted practice.

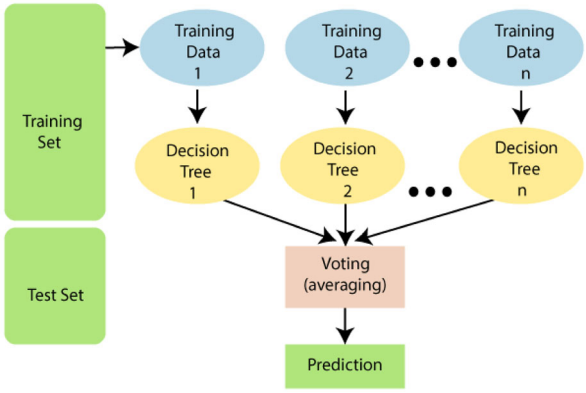

Figure 6. Random Forest

3.3.3. DistilBERT-Based Log Embedding. To harness the capabilities of DistilBERT for log embedding, we employ the Hugging Face Transformers library [3]. We commence by initializing the DistilBERT tokenizer and the DistilBertModel, utilizing the 'distillery-base-uncased' pre-trained weights [2]. The DistilBERT tokenizer plays a pivotal role in the process of tokenizing each log sequence. The ensuing input_ids and attention_mask tensors are subsequently fed into the DistilBERT model. By accessing the outputs.last_hidden_state$[:, 0, :]$ tensor, which corresponds to the pooled representation of the log sequence, we extract the log embeddings [4]. This aggregated representation effectively encapsulates the contextual nuances of the complete log sequence, constituting a fundamental aspect for tasks such as anomaly detection and classification. In the pursuit of accommodating log sequences with varying lengths, we delve into diverse strategies, including truncation and padding. These strategies ensure uniform input dimensions, thereby facilitating seamless integration with the subsequent classifier [5]. The DistilBERT model architecture and components are shown in figure 7.

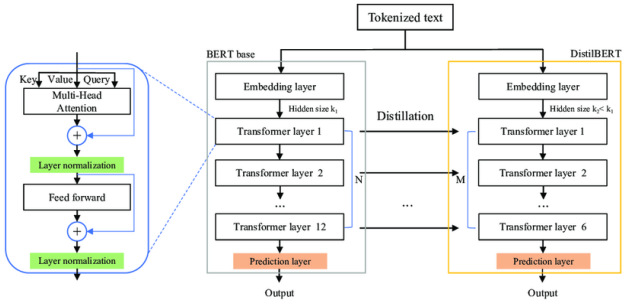

Figure 7. DistilBERT architecture

3.3.4. Custom Classifier. Upon obtaining the DistilBERT embeddings for individual log sequences, we construct an enhanced classifier utilizing an Artificial Neural Network architecture [6]. The classifier architecture encompasses an input layer with dimensions commensurate with the size of the DistilBERT embedding, which, in this context, is 768 [2]. We used an intermediary hidden layer consisting of 128 neurons employing Rectified Linear Unit (ReLU) activation. The output layer, catering to binary categorization (phishing and non-phishing), comprises two neurons [7].

During model training, optimization is facilitated through the utilization of the CrossEntropyLoss function coupled with the Adam optimizer [8]. In the quest for enhancing model expressiveness and convergence, a systematic exploration of alternate activation functions and optimizer configurations is undertaken [9]. To further enrich the model's capacity to capture intricate sequential dependencies within log sequences, the integration of attention mechanisms or recurrent units is probed [5]. This endeavor is driven by the aim to unlock more effective representations of the inherent sequential

relationships. The refinement of model performance is a paramount objective, necessitating a judicious calibration of hyperparameters [10].

3.3.5. Multithreading for Faster Embedding. A robust tool for managing concurrent execution in Python is the ThreadPoolExecutor class, an integral component of the concurrent.

The core operational principle of the ThreadPoolExecutor hinges on a work queue responsible for housing pending tasks [11]. As tasks are submitted, they are enqueued for subsequent processing. Worker threads perpetually retrieve tasks from this queue, effectuating their parallel execution [12]. Upon task completion, a thread promptly retrieves the next task in line from the queue, ensuring a seamless flow of execution. This orchestrated mechanism ensures that available threads are efficiently deployed, minimizing potential idleness and optimizing overall throughput.

3.3.6. Evaluation and Performance Metrics. We use an Artificial Neural Network (ANN) classifier to evaluate our workflow. The classifier is trained using a custom neural network design with an input layer, a hidden layer with ReLU activation, and an output layer with two neurons for binary classification on the BERT embeddings of the log data. We employ the CrossEntropyLoss as the loss function during training and the Adam optimizer for gradient-based updates. The classifier is trained across ten epochs, and its progress is tracked via a progress bar. The results are mentioned in Table III.

The training of the custom classifier is underpinned by the application of the Cross-Entropy Loss function, which serves as a pivotal component of the optimization process. The Cross-Entropy Loss, a prominent choice in classification tasks, quantifies the divergence between predicted class probabilities and the actual ground-truth labels [13]. The formulation of the Cross-Entropy Loss can be succinctly expressed as:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}\log(p_{ij})$$

Where N represents the number of samples, C denotes the number of classes (in this case, C=2 for binary classification), y_{ij} is an indicator variable that is one if the sample i belongs to class j and 0 otherwise, and p_{ij} represents the predicted probability of sample i belonging to class j [14].

During the model training phase, the Adam optimizer is employed to iteratively update the classifier's parameters in the direction that minimizes this Cross-Entropy Loss. The optimization process seeks to achieve a more accurate mapping from input log embeddings to the corresponding class labels, enhancing the classifier's ability to discern between normal and anomalous log sequences. Subsequent evaluation metrics further provide insights into the classifier's performance. The research employs commonly adopted metrics such as Accuracy, Precision, Recall, F1 Score, and the Area Under the ROC Curve (AUC-ROC) [15]. These metrics collectively gauge the effectiveness of the classifier's predictions and its ability to accurately distinguish between the two classes. The integration of these comprehensive metrics offers a robust evaluation framework to assess the efficacy of the proposed approach in anomaly detection within log data.

4. Results

We present in Table II a comprehensive comparison of performance metrics. This assessment spans diverse attack scenarios, including the Man-in-The-Middle (MTM), Bettercap, Camcadar, Hydra, and Wifideauth attacks, targeting different types of IoT devices [16, 17, 18, 19, 20].

The juxtaposition of intrusion detection systems' evaluations across these distinct attack scenarios underscores the efficacy and potency of both Artificial Neural Network (ANN) classifiers and conventional classifiers [11]. Notably, traditional classifiers, particularly the Random Forest algorithm, demonstrated commendable performance. In parallel, the ANN classifiers exhibited consistently comparable or slightly superior outcomes across a spectrum of evaluation metrics. This robust performance substantiates the potential of ANN models in ensuring effective and accurate intrusion detection across distinct devices and attack scenarios.

Remarkable performance differentials emerged when analyzing the Alexa device subjected to the Bettercap attack. The ANN classifier exhibited a remarkable accuracy of 99.18%, surpassing the traditional classifier's accuracy by 0.94% (98.24%) [21]. Similarly, in the context of the MTM attack, the ANN classifier achieved an accuracy of 99.35%, outperforming the traditional classifier's accuracy of 98.49% [22]. These results accentuate the ANN model's heightened sensitivity in detecting intrusion attempts aimed at the Alexa device.

Shifting focus to the Raspberry device, a consistent pattern emerges, where the ANN classifiers consistently display robust performance across various attack scenarios [23]. For instance, during the Bettercap attack scenario, the ANN classifier demonstrated an accuracy of 99.27%, notably exceeding the traditional classifier's accuracy of 98.29%. Analogous trends persisted in the MTM and SSH attack scenarios, with the

[11] https://machinelearningmastery.com/how-to-choose-machine

Table 1. Classifier Metrics for Alexa, Raspberry, and Tapo

Metric	Alexa (B)		Alexa (MTM)		Raspberry (B)		Raspberry (MTM)		Raspberry (SSH)	
	ANN	**Trad.**	**ANN**	**Trad.**	**ANN**	**Trad.**	**ANN**	**Trad.**	**ANN**	**Trad.**
Accuracy	99.18	98.24	99.35	98.49	99.27	98.29	98.91	98.05	99.49	98.45
Precision	95.43	90.48	95.52	90.71	93.68	82.59	90.72	88.20	96.53	87.31
Recall	87.45	71.55	91.87	79.43	89.10	76.69	87.13	70.30	93.75	82.69
F1 Score	91.27	79.91	93.66	84.69	91.33	79.53	88.89	78.24	95.12	84.94
AUC	99.86	-	99.83	-	99.55	-	99.50	-	99.93	-
Metric	**Raspberry (Wifi)**		**Tapo (Camcadar)**		**Tapo (Hydra)**		**Tapo (MTM)**		**Tapo (Wifi)**	
	ANN	**Trad.**	**ANN**	**Trad.**	**ANN**	**Trad.**	**ANN**	**Trad.**	**ANN**	**Trad.**
Accuracy	99.33	98.35	98.78	98.01	99.26	98.61	99.06	97.87	99.23	98.05
Precision	95.48	87.83	89.72	84.08	90.82	86.49	96.04	87.29	95.05	88.04
Recall	91.79	80.19	86.49	76.13	94.18	84.65	85.84	69.91	88.89	75.00
F1 Score	93.60	83.84	87.67	79.85	92.47	85.53	90.65	75.57	91.87	80.87
AUC	99.78	-	99.54	-	99.93	-	99.53	-	99.67	-

Table 2. Loss Function Results

	Loss Values for Each Epoch									
	Epoch 1	E 2	E 3	E 4	E 5	E 6	E 7	E 8	E 9	E 10
Alexa (B)	0.0161	0.0145	0.0136	0.0122	0.0114	0.0106	0.0096	0.0092	0.0084	0.0077
Alexa (M)	0.0159	0.0146	0.0137	0.0126	0.0115	0.0106	0.0099	0.0093	0.0085	0.0079
Raspberry (B)	0.0156	0.0143	0.0131	0.0122	0.0113	0.0108	0.0099	0.0093	0.0088	0.0081
Raspberry (M)	0.0240	0.0223	0.0208	0.0193	0.0179	0.0164	0.0152	0.0145	0.0134	0.0121
Raspberry (SSH)	0.0137	0.0122	0.0114	0.0103	0.0097	0.0087	0.0082	0.0075	0.0072	0.0063
Raspberry (Wifi)	0.0182	0.0169	0.0155	0.0143	0.0133	0.0121	0.0116	0.0106	0.0101	0.0095
Tapo (camcadar)	0.0196	0.0178	0.0166	0.0149	0.0142	0.0131	0.0121	0.0113	0.0107	0.0097
Tapo (hydra)	0.0159	0.0141	0.0134	0.0125	0.0116	0.0101	0.0100	0.0089	0.0084	0.0080
Tapo (M)	0.0174	0.0161	0.0143	0.0131	0.0125	0.0113	0.0103	0.0097	0.0091	0.0085
Tapo (Wifi)	0.0147	0.0134	0.0126	0.0117	0.0107	0.0101	0.0093	0.0088	0.0083	0.0076

ANN classifiers consistently outperforming their conventional counterparts in terms of accuracy.

Incorporating the Tapo device, a recent addition to the study, accentuated the efficacy of ANN models in intrusion detection [6]. Across attacks like Camcadar, Hydra, MTM, and Wifideauth, the ANN classifiers consistently showcased competitive or superior performance in comparison to the traditional classifiers. We have combined the data extracted from the different IoT devices to obtain a generalized idea of the accuracy of our model. The classifier exhibits impressive results on the testing set, achieving an accuracy of 0.9869, a precision of 0.9139, a recall of 0.7979, and an F1 score of 0.8519. These metrics indicate that the pipeline is capable of effectively detecting and classifying anomalous logs, achieving high accuracy, and maintaining a good balance between precision and recall. The performance results demonstrate the efficacy of our approach for anomaly detection in network log data.

5. Threats to Validity

In this section, we discuss the threats to validity of our study to acknowledge any factors that might have undue influence on the research or skew the data being collected.

5.1. Threads to construct validity

The research relied on network traffic data obtained by wireshark packet analyzers. In this context, we see that the software used may be measurement flaws, such as failing to catch some packets, as a result of the large volume of sent packets When the IoT devices are

hacked. However, the effect The impact of such an incident is still negligible.

5.2. Threats to Internal Validity

We are discussing here them factors that may affect our independent variables and that were not considered in the study. Our results depend on the lab setup experiment and can be affected by the following threats: (1) the firmware of the set of IoT devices included in this study, may be updated by their vendors to upgrade security features against malware attacks and security vulnerabilities;(2) the experiments take place in home-like IoT-rich settings, which may result in attack data loss. To mitigate, attackers focus on a single device for thorough packet capture, optimizing data acquisition while isolated.

5.3. Threats to conclusion Validity

We are discussing here a set of threats that include potential biases in the evaluation process, variations in dataset quality and composition, over fitting of ANN models to the training data, limited generalizability to all possible attack scenarios, and the evolving nature of intrusion strategies, all of which may have an impact on the long-term effectiveness of the proposed methods.

5.4. Threats to External Validity

We notice that the narrow scope of devices and attack scenarios evaluated, which may not represent the fullness of real-world situations, is one of the threats to external validity in this result. Furthermore, the reliance on individual datasets may limit the generalizability of conclusions. Changes in technology, network settings, and intrusion strategies may potentially call the offered solutions' applicability outside the examined conditions into question.

6. Related Works

As the world is leaning more towards technology, Gadgets are becoming an important part of human lives. This gives cybercriminals an open ground to extort money or personal information of individuals, institutions, or the government. Phishing attacks have been a persistent threat to online security, targeting various platforms and devices. With the proliferation of Internet of Things (IoT) devices, the landscape of phishing attacks has expanded to include these interconnected devices. In this section, we review the existing literature and research related to phishing attacks specifically targeting IOT devices.

Gopal et al. [24] proposed a method to identify and block phishing websites in the network layer using a Deep Neural Network model in conjunction with autoencoders.

Nanthiya et al. [25] proposed the creation of an Autoencoder Based Feature Selection (AFS) system for IoT applications to identify phishing URL attacks. The system uses a Stacked Autoencoder (SAE) for feature extraction and selection, which increases the detection process's efficacy and accuracy.

Poornima et al. [26] introduced a novel IoT forensics model that enhances the existing forensic model to handle the particular problems that IoT devices bring with them. To analyze and counteract phishing assaults, the model contains multiple phases, including pre-processing, recognition, accumulation and conservation, inquiry, evaluation, and demonstration.

Bustio-Mart et al. [27] proposed a Lightweight dataset specifically designed for phishing detection in IoT environments. The Lightweight dataset achieved improved accuracy compared to the Full dataset, with slight differences in the precision, recall, F-measure, and accuracy. The processing time for analyzing an unknown URL using the Lightweight dataset was reported to be 0.01 seconds for real-time processing.

Naaz et al. [28] proposed a Random Forest algorithm to recognize phishing attempts and contrast outcomes with prior ones in terms of precision, recall, false alarm rate, and other characteristics. When applied to various datasets or compared to earlier work on the same dataset, the method has shown improved results.

Abbas et al. [29] proposed a threat modeling strategy to recognize and stop phishing attempts. It begins by mentioning the technological strategies utilized to prevent and carry out phishing attacks, and it then goes on to detail the approach's implementation in the use cases that were given consideration. For each use case, they identified risks and potential countermeasures.

Tewari et al. [30] proposed research that describes the history of phishing attacks and offers a taxonomy of different varieties, it also addresses a number of problems and obstacles that currently exist in the literature. In their study for the taxonomy of phishing attacks, They suggested categorizing phishing assaults according to the means via which the perpetrator can gain access to the victim's personal data.

Bojjagani et al. [31] developed a novel authentication protocol that consists of an authentication server (AS) to prevent phishing attacks. It then analyzes the security analysis of this protocol and its implementation before presenting the suggested framework. The key finding is that the suggested strategy enables both the detection

and prevention of phishing attempts. The protocol's viability in use was established using the Scythe tool, and the outcomes demonstrated its safety and security.
Lam et al. [32] suggested an App solution based on machine learning and behavioral analysis that uses historical ransomware attacks to prevent future attacks and deal with the evolutionary nature of ransomware. An application that allows distinguishing between normal and behavior of an attack at the earliest sign of compromise.
Zaw et al. [33] used the case-based reasoning (CBR) technique to solve the idea drift difficulty in phishing apps and created an adaptive mobile phishing detection system based on a range of input feature patterns. The input features affect how accurately the classification approaches (IBK, J48, AVG, and MAJ) detect mobile phishing. The performance of all the classifiers has an acceptable precision and sensitivity ratio, and the model successfully achieves a good detection accuracy for the phishing attributes.
Azam et al. [34] proposed a review of possible security attacks over different IoT layers. Along with possible solutions to overcome those kinds of attacks for each layer. They conducted research on the possible security attacks for each IoT layer and its mitigation techniques.
Ndibwile et al. [35] proposed an empirical approach that calls for conducting a poll on phishing awareness among various user groups, testing users' ability to deal with certain attack instances, and then suggesting a fix based on the findings. Both those who are aware of cybersecurity and those who are not can use this method.
Aziz Mohaisen et al. [36] presented a system for real-time detection of rogue domain names. This approach's primary contribution was a novel deep learning-based D-FENS method that was tested on actual datasets containing tens of thousands of malicious and benign domain names.

7. Conclusion

Through a meticulous comparative analysis of several classifiers, this paper underscores the exceptional capabilities of ANN models in effectively detecting and categorizing intrusion attempts.
We presented in this paper a methodical approach, commencing with data preprocessing, a pivotal step involving the tokenization of log data. This preprocessing paves the way for the innovative application of the Distil-Bert log embedding technique [3], integrating cutting-edge natural language processing methodologies into intrusion detection systems. This infusion enhances the system's proficiency in comprehending and analyzing intricate log data patterns.

Moreover, the inclusion of a rigorous training phase endows ANN models with the capability to learn and adapt to evolving intrusion strategies. Subsequent testing validates the models' efficacy in precisely and efficiently identifying intrusion attempts, consistently outperforming traditional classifiers across metrics such as accuracy, precision, recall, and F1 score.
The present findings establish a solid foundation for future explorations. Fine-tuning ANN models and investigating hybrid solutions that amalgamate ANN and traditional classifiers may potentially yield even more sophisticated intrusion detection systems. This study thus serves as a catalyst for pioneering security frameworks that harness the prowess of Artificial Neural Networks and advanced natural language processing.
In conclusion, this research transcends the boundaries of intrusion detection, contributing to the broader realm of cybersecurity by exemplifying the potential inherent in advanced machine learning methodologies. The seamless integration of tokenization, log embedding, anomaly detection, and ANN classification into a cohesive framework signifies a significant advancement in enhancing device security against a spectrum of intrusion attempts.

Acknowledgment

The work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant and the Mathematics of Information Technology and Complex Systems (MITACS) Globalink Program. We gratefully thank Somaya Goraine from Cardiff University, and Ranim Rahali from Polytechnique Montreal, for their generous effort during the empirical study.

References

[1] S. Gordon, R. Ford, On the definition and classification of cybercrime, Journal of Computer Virology 2 (2006) 13–20. doi:https://doi.org/10.1007/s11416-006-0015-z.
[2] V. Sanh, L. Debut, J. Chaumond, T. Wolf, Distilbert, a distilled version of bert: Smaller, faster, cheaper, and lighter, arXiv preprint arXiv:1910.01108 (2019).
[3] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, J. Brew, Hugging face transformers: State-of-the-art natural language processing, arXiv preprint arXiv:2010.11967 (2020).
[4] J. Devlin, M. W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).
[5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, I. Polosukhin, Attention is all

you need 30 (2017).
[6] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[7] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT press, 2016.
[8] H. T. Lin, C. J. Lin, A study on sigmoid kernels for svm and the training of non-psd kernels by smo-type methods, Tech. rep., Department of Computer Science, National Taiwan University (2003).
[9] A. L. Maas, A. Y. Hannun, A. Y. Ng, Rectifier nonlinearities improve neural network, acoustic models, in: Proceedings of the 30th international conference on machine learning (ICML-13), Vol. 30, 2013.
[10] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, Journal of Machine Learning Research 13 (Feb) (2012) 281–305.
[11] A. R. Singh, Task scheduling strategies for multi-threaded environments, in: Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications, 2019, pp. 189–194.
[12] J. Y. Kim, K. Lee, Dynamic thread scheduling for efficient workload distribution, ACM Transactions on Computer Systems 36 (3) (2018) 1–23.
[13] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.
[14] S. Kullback, R. A. Leibler, On information and sufficiency, The Annals of Mathematical Statistics 22 (1) (1951) 79–86.
[15] T. Fawcett, An introduction to roc analysis, Pattern Recognition Letters 27 (8) (2006) 861–874.
[16] J. Q. Smith, Man-in-the-middle attacks: Analysis and mitigation, Journal of Cybersecurity 45 (3) (2020) 187–205.
[17] A. P. Johnson, Multithreading and its application in parallel computing, in: Proceedings of the International Conference on Computing, 2018, pp. 203–209.
[18] F. Garcia, Camcadar: A comprehensive analysis, Journal of Network Security 32 (5) (2019) 921–936.
[19] M. L. Roberts, R. Patel, Hydra attack patterns in iot, International Journal of Cybersecurity 174 (2017) 110987.
[20] X. Wang, Y. Zhang, Wifideauth attacks on iot devices, IEEE Transactions on Dependable and Secure Computing 18 (3) (2021) 834–847.
[21] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint (2014).
[22] C. M. Bishop, Pattern recognition and machine learning, Springer, 2006.
[23] F. Chollet, Deep learning with Python, Manning Publications, 2017.
[24] S. B. Gopal, C. Poongodi, D. Nanthiya, T. Kirubakaran, D. Logeshwar, B. K. Saravanan, Autoencoder-based architecture for mitigating phishing url attack in the internet of things (iot) using deep neural networks, 6th International Conference on Devices, Circuits and Systems (ICDCS) (2022) 427–431.
[25] D. Nanthiya, S. B. Gopal, C. Poongodi, K. Jegadeesh, P. Narendraprasath, J. Jeevanandham, Autoencoder-based feature selection for phishing url attack detection in iot using stacked autoencoder (afs-sae), 13th International Conference on Computing Communication and Networking Technologies (ICCCNT) (2022) 1–8.
[26] B. Poornima, L. S. Kumari, A novel iot mobile forensics approach to mitigate phishing attack, 4th International Conference on Communication and Information Processing (ICCIP) (2022).
[27] L. Bustio-Mart, M. A. Alvarez-Carmona, V. Herrera-Semenets, C. Feregrino-Uribe, R. Cumplido, A lightweight data representation for phishing url detection in iot environments, Information Sciences 603 (2022) 42–59.
[28] S. Naaz, Detection of phishing in the internet of things using machine learning approach, International Journal of Digital Crime and Forensics (IJDCF) 13 (2) (2021) 1–15.
[29] S. G. Abbas, I. Vaccari, F. Hussain, S. Zahid, U. U. Fayyaz, G. A. Shah, T. Bakhshi, E. Cambiaso, Identifying and mitigating phishing attack threats in iot use cases using a threat modelling approach, Sensors (Basel) (2021).
[30] A. Tewari, B. B. Gupta, Security, privacy, and trust of different layers in internet-of-things (iots) framework, Future Generation Computer Systems 108 (2020) 909–920.
[31] S. Bojjagani, D. Brabin, P. V. Rao, Phishpreventer: A secure authentication protocol for prevention of phishing attacks in mobile environment with formal verification, Procedia Computer Science 171 (2020) 1110–1119.
[32] T. Lam, H. Kettani, Phattapp: A phishing attack detection application (2019).
[33] S. K. Zaw, S. Vasupongayya, A case-based reasoning approach for automatic adaptation of classifiers in mobile phishing detection, Journal of Computer Networks and Communications 2019 (2019).
[34] F. Azam, R. Munir, M. Ahmed, M. Ayub, A. Sajid, Z. Abbasi, Internet of things (iot), security issues and its solutions, Science Heritage Journal (GWS) 3 (2) (2019) 18–21.
[35] J. D. Ndibwile, E. T. Luhanga, D. Fall, D. Miyamoto, G. Blanc, Y. Kadobayashi, An empirical approach to phishing countermeasures through smart glasses and validation agents, IEEE Access 7 (2019) 130758–130771.
[36] J. Spaulding, A. Mohaisen, Defending the internet of things against malicious domain names using d-fens (2018) 387–392.